

THE INFLUENCE OF LANGUAGE AND CULTURE ON THE UNDERSTANDING OF VOCAL EMOTIONS

Rene Altrov and Hille Pajupuu
Institute of the Estonian Language

Abstract. We investigated the influence of culture and language on the understanding of speech emotions. Listeners from different cultures and language families had to recognize moderately expressed vocal emotions (joy, anger, sadness) and neutrality of each sentence in foreign speech without seeing the speaker. The web-based listening test consisted of 35 context-free sentences drawn from the Estonian Emotional Speech Corpus. Eight adult groups participated, comprising: 30 Estonians; 17 Latvians; 16 North-Italians; 18 Finns; 16 Swedes; 16 Danes; 16 Norwegians; 16 Russians. All participants lived in their home countries and, except the Estonians, had no knowledge of Estonian. Results showed that most of the test groups differed significantly from Estonians in the recognition of most emotions. Only Estonian sadness was recognized well by all test groups. Results indicated that genealogical relation of languages and similarity of cultural values are not advantages in recognizing vocal emotions expressed in a different culture and language.

Keywords: emotions, emotional speech, vocal expression, perception, cross-cultural, genealogically related languages, European languages

DOI: <http://dx.doi.org/10.12697/jeful.2015.6.3.01>

1. Introduction

Interpersonal communication involves various non-verbal skills, including the use of the surrounding space, gestures and facial expressions, postures, the way of keeping eye contact, and so forth. A successful conversation largely depends on how we express our own emotions, how we understand those of others and how adequate our reaction is to their emotions. Thus, emotions have a central role in our lives, being present in most everyday communication (see Cowie et al. 2011).

Generally, we can understand a person's emotions by seeing and hearing them. If we sit in a cafe in a foreign country while someone sitting at the next table raises their voice waving their hands about,

and maybe even jumps up and slams the door behind them, we hardly need to know the language to guess that the customer is outraged. Such full-blown emotions, which are accompanied by physical expression (e.g. becoming red in the face, panting or bursting into tears) are easy to recognize. Such emotions can be regarded as universal inasmuch as they can be understood whatever the language, culture or personality.

Understanding becomes more complicated in the case of moderately expressed emotions, which are prevalent in everyday communication. For example, everyday joy is a pleasant rather than exalted feeling, anger is hardly expressed as rage, but rather as irony or resentment, and sadness is usually expressed as anxiety rather than despair. Therefore we can say that universality applies only to the classification of emotions into positive and negative, whereas the expression and understanding of emotions depend on culture and language (Altrov 2013, Altrov and Pajupuu 2010, Elfenbein 2013, Kamaruddin et al. 2012, and Pell et al. 2009).

Cross-cultural research on the recognition of emotions from oral speech has provided evidence for universality as well as for cultural specificity (e.g. Altrov 2013, Bryant and Barrett 2008, and Jürgens et al. 2013). According to Laukka et al. (2014) as well as to Scherer et al. (2011) the universality of emotions is supported by the fact that if a listening task requires the classification of emotions of any culture into predetermined categories, recognition accuracy is better than chance probability. However, research has revealed that emotion expressions are better identified by members of one's own culture (so-called in-group advantage). A poorer performance in identifying the emotion expressions of other cultures may be due to cultural differences in emotion expression (Elfenbein 2013, and Laukka et al. 2014). True, most proof of the universality of emotion recognition has been obtained from cross-cultural studies of the face, while cross-cultural studies of the identification of vocal emotions in speech are much rarer (see Paulmann and Uskul 2014, and Pell et al. 2009).

Paulmann and Uskul (2014) have provided a three-way division of the design of studies addressing the recognition of vocal emotions across different culture or language groups: (1) one against all; (2) all against one; (3) all against all. Despite the differences in research design, the aim of all those studies is to find out to what extent emotion recognition is universal or determined by language and culture. In the case of *one against all* listeners from different cultures are asked to identify emotions communicated by speakers from another culture.

The group can be exemplified by a study by Scherer et al. (2001), in which listeners from nine countries listened to 30 pseudo-sentences presented with five emotions by four native German actors. All test groups recognized fearful, joyful, sad, angry, and neutral sentences better than chance probability, but German listeners performed significantly better than the rest. It was also observed that emotion recognition was supported by linguistic similarity to the test language. Thus German emotions were identified better by the Dutch than the Malay group, whose language is very different from German.

According to the results of Altrov (2013), emotion identification involves both universal and sociocultural factors. Her listening test required that three groups of participants – native Estonian speakers living in Estonia, native Russian speakers living in Estonia who can speak Estonian, and native Russian speakers living in Russia who cannot speak Estonian – listen to Estonian sentences classifiable by emotions. Four emotion classes – anger, joy, sadness, and neutrality – were involved. Native Russian speakers living in Estonia identified all emotions with an accuracy of over 50%. The performance of native Russian speakers living in Russia was less perfect – they gained scores of over 50% only for Estonian sadness. Thus, the Russians living in Estonia recognized Estonian emotions to an extent that was rather more like the local Estonians than the Russians living in Russia. Hence Altrov (2013) concluded that joy, anger and neutrality are expressed in a culture-specific manner, while the expression of sadness could be universal.

The study of Jürgens et al. (2013) was aimed at discovering the possible cultural influence on emotion understanding in German play-acted and spontaneous speech. Both the spontaneous and play-acted expressions of emotions (anger, fear, joy, and sadness) were presented for listening to three groups – native Germans, Romanians, and Indonesians. The latter two were chosen as representatives of a collectivist culture (see Hofstede 2001) versus the individualistic German culture. Taking into account the controversial results of earlier studies as to whether collectivist cultures are better at recognizing negative emotions, the authors explored this as well. It turned out that although the emotion recognition rates were low in all three groups, the Germans scored slightly higher, which indicates a moderate in-group advantage. The hypothesis that collectivist cultures identify negative emotions better than individualistic ones was partially confirmed; notably, the Indonesians and Romanians recognized play-acted anger and spontaneous

sadness. No cultural difference was revealed in the recognition of fear. The results support the view that emotion expression has universal as well as culturally specific aspects.

If the research is designed as *all against one*, listeners from one culture are asked to identify emotions expressed by speakers from different cultural groups. For example Pell et al. (2009) investigated how well 61 adult native speakers of Argentine Spanish could identify five basic emotions (anger, disgust, fear, sadness, joy), pleasant surprise and neutral in pseudo-utterances play-acted by native speakers of Arabic, English, German and Spanish. According to their results, the listeners were able to identify the emotion of the pseudo-utterances whoever the speaker, but nevertheless the rate of identification was higher if the listeners shared the speaker's native language. In addition, it was revealed that whatever the native language of the speaker, emotion recognition was higher for anger and sadness, and the lowest for disgust. The only emotion revealing a distinct difference between a native and a foreign speaker was joy. Notably, the native speakers of Argentine Spanish were significantly better at recognizing the joy expressed by native Argentine speakers, than by speakers of all other languages. Another interesting point of the study is how the native speakers of Argentine Spanish, whose language belongs to the Romanic branch of the Indo-European language family, managed to recognize emotions expressed by speakers from a different branch of the same family (English and German) and from the Semitic branch of the Afro-Asiatic language family (Arabic). Even though Arabic was the most dissimilar of the languages involved, the difference did not seem to affect the understanding of the emotions rendered by Arabic pseudo-utterances. This outcome suggests that linguistic similarity is not a consistent factor in predicting the accuracy of vocal emotion recognition across languages.

Thompson and Balkwill (2006) also demonstrated that understanding of emotions is affected by both universal and culturally specific factors. The study required 20 speakers of English to listen to emotion sentences with a neutral semantic content, expressing joy, sadness, anger and fear, as presented by male and female speakers in English, Chinese, Japanese and Tagalog. Recognition accuracy was better than chance probability for all emotions in all languages, while anger and sadness were recognized better across languages than joy and fear. The highest recognition rates were scored for English sentences and the lowest for Japanese and Chinese ones. Hence the researchers concluded that the listeners'

ability to recognize emotions in totally unfamiliar languages suggests that some of the prosodic characteristics of vocal emotions may indeed be universal.

In the case of a fully crossed design, *all against all*, listeners from different cultures are asked to identify emotions expressed by each other (see Paulmann and Uskul 2014). Research by Paulmann and Uskul (2014) is a good example of the design requiring mutual recognition of seven emotions (anger, disgust, fear, happy, sad, surprise, neutral) as expressed in pseudo-sentences by British and Chinese speakers. According to the results, both groups were able to recognize emotions in their native language as well as in a foreign language, with a recognition accuracy better than chance probability. However, members of both cultural groups were more accurate in recognizing the emotions expressed by a member of their own cultural group than by one of the other cultural group. The British participants manifested in-group advantage in recognizing all emotions except joy and neutrality; the Chinese participants showed in-group advantage in recognizing all emotions except fear and sadness.

Laukka et al. (2014) used a machine learning method to classify, on the basis of acoustic features, eleven emotions expressed with a different intensity in a sentence with a constant verbal content by 100 actors from five English-speaking countries. Experiments were performed in conditions where classifier programmes were trained on stimuli from either the same or different culture vis-à-vis the stimuli subsequently used in the testing phase. If, in cross-cultural conditions, a particular emotion was classified with an accuracy above chance, it was considered indicative of the fact that the emotion is expressed acoustically similarly across all cultures involved. A wide range of emotions were classified with an accuracy better than chance probability. The largest differences were revealed for pride, anger, sadness and fear, while the smallest were observed for happiness, lust and relief. However, classification accuracy was higher in intra-cultural versus cross-cultural conditions, which was suggestive of in-group advantage. Also, it seemed possible that cultures with a similar cultural profile according to the cultural dimensions described by Hofstede (2001), such as, for example, Australia and USA, express emotions in a similar way.

The described studies offer proof that both expression and understanding of emotions involve universal aspects as well as culturally and linguistically specific ones. Several studies also point out the role of in-group advantage in emotion recognition, where linguistic as well as cultural similarity may produce a supportive effect.

In the current work, we aimed to add to the above knowledge by providing information gained of some languages and cultures not covered so far, in order to learn to what extent linguistic and cultural affiliation may affect the understanding of vocal emotions. To achieve this aim, we decided to investigate how non-speakers of Estonian who do not live in Estonia and who belong to different language branches and different cultures can understand moderately expressed elicited emotions in Estonian speech. The participants were native speakers of the titular language from Estonia, Latvia, Finland, Sweden, Denmark, Norway, Italy and Russia. In addition to linguistic differences the countries differed in the cultural dimensions according to Hofstede (2001).

As listeners from different cultures were asked to recognize the vocal emotions expressed by a native Estonian speaker, who represented still another culture, our research design can be classified as *one against all*, in terms of Paulmann and Uskul (2014).

We hypothesized:

H1 (hypothesis of universality): Cultures that speak different languages and have different cultural values will be able to recognize vocal emotions with an accuracy better than chance probability, without confusing them with other emotions.

H2 (hypothesis of similarity): Cultures with similar cultural values and of the same language branch perform similarly in emotion recognition.

2. Method

2.1. Material

The material analysed comes from the Estonian Emotional Speech Corpus.¹ The corpus contains 1,234 sentences read by a female voice. These sentences (all different) have been extracted from longer passages of recorded text. The reader has not been instructed to use any particular emotion while reading the passages, assuming that any text evokes a certain mood, which sounds in the reading voice. To provide the sentences with an emotion label the sentences were extracted from the context and presented to a group of evaluators whose native language

1 <http://peeter.eki.ee:5000/>

was Estonian. They were asked to decide whether the sentence sounded joyful, angry, sad, or neutral (see Altrov and Pajupuu 2012).

For the present study the corpus was searched for 10 joyful, 10 sad, 10 angry and 5 neutral sentences with an emotion identification rate of no less than 65% (i.e. more than 2.5 times better than chance), where the semantic content had no particular role in emotion identification (see Table 1). The sentences were arranged into a web-based listening test.

Table 1. Emotional corpus sentences selected for the listening test (see Altrov 2013)

Corpus sentences selected for listening test (translated into English)	Recognition rate from listening (corpus data)
1. <i>Although Ott knew nothing of my existence ...</i>	joy 86.2%
2. <i>As I see it, the parliament acts as a “rubber stamp”.</i>	anger 75.8%
3. <i>Musicians and artists seem to enjoy interacting with their fellow unshaved citizens.</i>	anger 77.4%
4. <i>Our family never discussed any problems.</i>	sadness 82.8%
5. <i>I saw how the district chief of Nõmme passed out at the Raba race yesterday.</i>	neutral 71.0%
6. <i>Once, I wanted hot semolina porridge at three o'clock in the morning.</i>	joy 69.7%
7. <i>I live in Tallinn now, together with my invalid daughter, and we feel much more comfortable here than in Narva.</i>	neutral 65.5%
8. <i>However, the Estonian national team deserves praise.</i>	joy 87.1%
9. <i>At yesterday's concert Padar smoked like a chimney.</i>	neutral 66.7%
10. <i>I can't move my cowshed to the vicinity of Tallinn, can I?</i>	sadness 74.2%
11. <i>His spirit will have a most prolonged influence on us.</i>	sadness 93.1%
12. <i>Only state officials can expect a 10% rise.</i>	anger 75.9%
13. <i>She even called my granny, urging me to come back to school.</i>	joy 81.2%
14. <i>There could be a medical professional or two available by the wood trail, couldn't there?</i>	anger 90.3%
15. <i>What I may and what I may not.</i>	sadness 80.6%

Table 1. Continuation

Corpus sentences selected for listening test (translated into English)	Recognition rate from listening (corpus data)
16. <i>The more painful the truth will be later.</i>	neutral 75.0%
17. <i>A waiter said in the newspaper that his constantly bad mood was due to the numerous Finnish clientele.</i>	neutral 70.0%
18. <i>In this case, it is a beautiful suffering.</i>	joy 75.0%
19. <i>I can't see why people would want to look so much uglier than they really are.</i>	anger 96.7%
20. <i>Can't we really manage anything without training any more?</i>	anger 67.7%
21. <i>At that moment he was the only person there for me 24/7.</i>	sadness 94.8%
22. <i>So I quitted work without notice.</i>	sadness 90.6%
23. <i>At three in the morning!</i>	joy 75.0%
24. <i>If a man works at a restaurant, he is usually believed to have something wrong with him.</i>	anger 66.7%
25. <i>It would be so genuine!</i>	joy 74.2%
26. <i>Whatever I do, he is never happy.</i>	sadness 78.1%
27. <i>And dizzy like sharks.</i>	anger 65.6%
28. <i>I miss Enn even in the daytime when I work.</i>	joy 67.7%
29. <i>This is an enormous hole.</i>	sadness 75.9%
30. <i>He practically doesn't touch alcohol.</i>	joy 77.4%
31. <i>What those four years have done ...</i>	sadness 77.4%
32. <i>Like a dead body.</i>	anger 79.3%
33. <i>How to cope with such a situation?</i>	anger 83.9%
34. <i>But just a little bit.</i>	joy 78.1%
35. <i>A difficult passage.</i>	sadness 86.2%

According to the results of a statistical acoustic analysis of the sentences drawn from the Estonian Emotional Speech Corpus, the emotions manifested in Estonian read-out speech can be described by the following characteristics (Tamuri 2012, 2015, and Tamuri and Mihkla 2012):

Joy has a high pitch, average intensity and average speech rate.

Anger has a low pitch, average intensity and high speech rate.

Sadness has an average pitch, low intensity and slow speech rate.

Neutral speech has an average pitch, high intensity and average speech rate.

Parameter dynamics:

The mean intensity from strongest to weakest: neutral > anger > joy > sadness.

The mean pitch from highest to lowest: joy > neutral = sadness > anger.

The average speech rates from the most rapid to the slowest: anger > joy > neutral > sadness.

2.2. Participants

The listening test was taken by eight groups of adults:

30 Estonians (15 female, 15 male, aged 30–60, $M_{\text{age}} = 40.8$, $SD = 8.7$), mother tongue Estonian;

17 Latvians (8 female, 9 male, aged 30–55, $M_{\text{age}} = 33.1$, $SD = 6.6$), mother tongue Latvian;

16 North-Italians (8 female, 8 male, aged 30–66, $M_{\text{age}} = 42.8$, $SD = 10.5$), mother tongue Italian;

18 Finns (9 female, 9 male, aged 30–76, $M_{\text{age}} = 49.2$, $SD = 16.1$), mother tongue Finnish;

16 Swedes (8 female, 8 male, aged 30–56, $M_{\text{age}} = 45.2$, $SD = 6.8$), mother tongue Swedish;

16 Danes (8 female, 8 male, aged 30–66, $M_{\text{age}} = 45.4$, $SD = 9.7$), mother tongue Danish;

16 Norwegians (8 female, 8 male, aged 30–65, $M_{\text{age}} = 42.1$, $SD = 10.6$), mother tongue Norwegian;

16 Russians (8 female, 8 male, aged 30–41, $M_{\text{age}} = 33.8$, $SD = 4.1$), mother tongue Russian.

All the participants lived in their homelands and, except the Estonians, had no knowledge of Estonian. For the homelands of the participants see Figure 1.



Figure 1. Homelands of the participants.

The participants represent different language groups: Estonian and Finnish belong to the Finnic branch of the Uralic language family; Latvian, Italian, Russian, Swedish, Danish and Norwegian belong to the Indo-European language family, of which Latvian represents the Baltic branch, Italian the Romanic branch, Swedish, Danish and Norwegian the North Germanic branch and Russian the East Slavic branch.

If we compare the cultural values of the participants based on Hofstede's cultural dimensions (individualism vs. collectivism, uncertainty avoidance, power distance, and masculinity vs. femininity), we can see that their home countries are culturally different. The cultural indices for Estonia, Latvia, Italy, Finland, Sweden, Denmark, Norway, and Russia are given in Table 2.

Table 2. Cultural indices for Estonia, Latvia, Italy, Finland, Sweden, Denmark, Norway, Russia (Hofstede 2001 and Hofstede et al. 2010)

States	Individualism	Power distance	Uncertainty avoidance	Masculinity
Estonia (EE)	60	40	60	30
Latvia (LV)	70	44	63	9
Italy (IT)	76	50	75	70
Finland (FI)	63	32	59	26
Sweden (SE)	71	31	29	5

2.3. Procedure

On entering the web-based testing environment the participants were asked to fill in the following personal data: sex, age, education, ethnicity, mother tongue and main language of education. Instructions were available as soon as the test was opened. Next, they were asked to listen to context-free Estonian sentences and decide, without seeing the text, upon the emotion of the sentence. The choice was between joy, anger, sadness and a neutral attitude. It was explained that in normal speech, these emotions are seldom encountered in their full form and this is why joy should be interpreted as an emotion that also covers gratitude, happiness, pleasure and exhilaration; anger also includes resentment, irony, reluctance, contempt, malice and rage; and sadness covers loneliness, disconsolation, concern and hopelessness, while neutral refers to normal speech without special emotions. The participants could listen to each sentence as many times as they wished.

Pearson's chi-square test of independence was applied to evaluate the difference between the groups' abilities in emotion recognition (the difference is significant if $p < .05$). Because in some cases the cell count was below 5, Pearson's chi-square test was supplemented by Fisher's exact test, which allows for smaller cell counts. The results being similar, this article will present only the chi-square test results. A confusion pattern is also presented to demonstrate the accuracy of the recognition of the target emotion in groups and with what emotions the target emotion was confused. The target emotion is regarded as recognized correctly only if it has not been confused with any other emotion (no other emotion has probability exceeding chance).

3. Results

A confusion matrix was generated to demonstrate how the test groups' judgements were distributed between correct and incorrect responses (see Table 4).

All the test groups recognized the target emotion of the test sentences with an accuracy of more than chance probability. Yet only Estonians were consistent in the recognition of Estonian vocal emotions. For Estonians, no wrong choice was better recognized than chance probability, but part of the Latvian, Swedish and Norwegian participants mistook Estonian anger and joy for neutral; the Italians confused Estonian anger with neutral and Estonian neutral with joy; the Finns confused Estonian

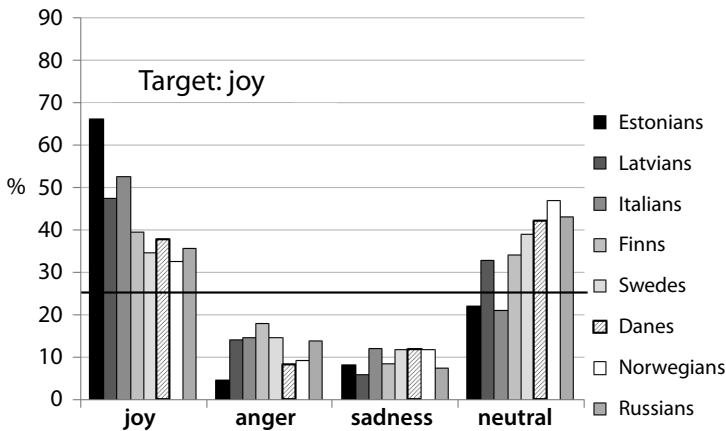
joy and anger with neutral and Estonian neutral with joy; the Russians confused Estonian joy and anger with neutral and Estonian neutral with anger. Only Estonian sadness was not confused with any other emotion and its accuracy of recognition was twice as high as chance probability. See Figures 2, 3, 4 and 5.

Table 4. Confusion matrix: Recognition of Estonian emotions by Estonians, Latvians, Italians, Finns, Swedes, Danes, Norwegians and Russians

Target emotion	Response emotions				Number of responses
	Joy (%)	Anger (%)	Sadness (%)	Neutral (%)	
<i>Estonians</i>					
Joy	65.9	4.3	8.0	21.7	299
Anger	2.7	57.0	16.8	23.5	298
Sadness	1.3	12.5	78.8	7.4	297
Neutral	6.7	10.7	11.3	71.3	150
<i>Latvians</i>					
Joy	47.4	14.1	5.8	32.7	156
Anger	9.8	39.2	17.0	34.0	153
Sadness	1.9	7.7	73.7	16.7	156
Neutral	11.3	22.5	13.8	52.5	80
<i>Italians</i>					
Joy	52.5	14.6	12.0	20.9	158
Anger	12.1	51.0	8.9	28.0	157
Sadness	1.9	6.9	82.4	8.8	159
Neutral	29.1	15.2	6.3	49.4	79
<i>Finns</i>					
Joy	39.5	18.0	8.4	34.1	167
Anger	14.6	42.7	10.4	32.3	164
Sadness	5.4	15.0	64.7	15.0	167
Neutral	34.5	15.5	4.8	45.2	84
<i>Swedes</i>					
Joy	34.6	14.5	11.9	39.0	159
Anger	17.6	32.7	10.7	39.0	159
Sadness	1.9	1.9	86.3	10.0	160
Neutral	5.1	20.5	14.1	60.3	78

Table 4. Continuation

Target emotion	Response emotions				Number of responses
	Joy (%)	Anger (%)	Sadness (%)	Neutral (%)	
Danes					
Joy	37.7	8.2	11.9	42.1	159
Anger	20.1	34.0	6.9	39.0	159
Sadness	0.6	5.7	79.2	14.5	159
Neutral	8.8	15.0	20.0	56.3	80
Norwegians					
Joy	32.5	9.1	11.7	46.8	154
Anger	10.5	36.8	13.8	38.8	152
Sadness	4.6	7.8	77.1	10.5	153
Neutral	7.6	24.1	10.1	58.2	80
Russians					
Joy	35.6	13.8	7.5	43.1	160
Anger	10.0	42.5	13.1	34.4	160
Sadness	6.3	6.9	73.1	13.7	160
Neutral	18.8	30.0	7.5	43.7	80

**Figure 2.** Recognition of JOY. *The black line indicates chance probability.*

Most groups confused Estonian joy with neutral. Only Italians did not confuse it with neutral any more than did the Estonians.

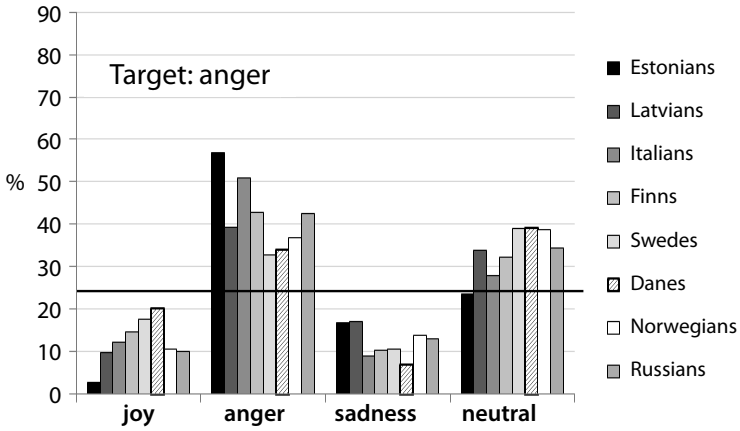


Figure 3. Recognition of ANGER. *The black line indicates chance probability.*

All groups except the Estonians confused Estonian anger with neutral.

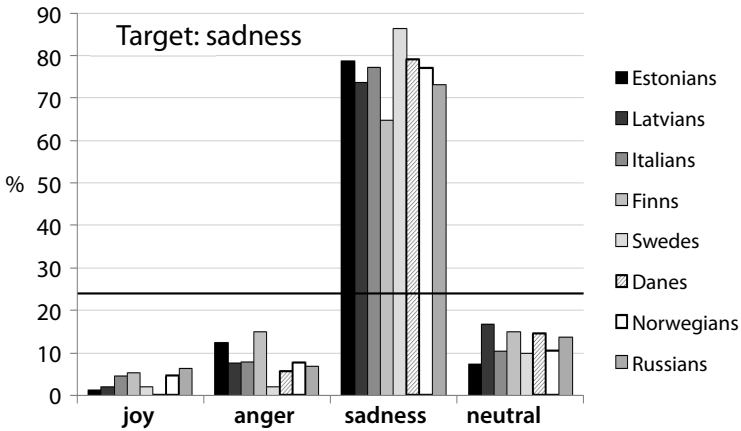


Figure 4. Recognition of SADNESS. *The black line indicates chance probability.*

Estonian sadness was not confused either with any other emotions or neutral.

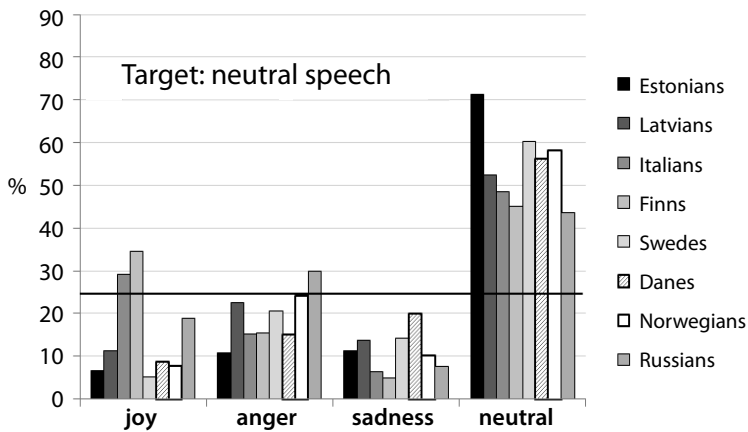


Figure 5. Recognition of NEUTRAL SPEECH. *The black line indicates chance probability.*

The Italians and the Finns confused neutral speech with joy, the Russian group confused it with anger.

Pearson's chi-square test was used to find out whether the eight groups of participants – Estonians, Latvians, Italians, Finns, Swedes, Danes, Norwegians and Russians – significantly differed from each other in terms of emotion recognition (considering both correct and incorrect responses).² Tables 5–8 present a pairwise comparison of the groups in recognizing Estonian vocal emotions.

² The null hypothesis is that there is no significant difference between the variables. The results are then given a level of significance. If this significance is low enough the null hypothesis can be neglected and a difference is established. The level of significance had to be below .05 to be regarded as showing a significant difference between the variables.

Table 5. Target emotion JOY. Comparison of test groups

Target Groups	Joy	Anger	Sadness	Neutral	p	Pearson's chi-square statistic as distance measure	
JOY	EE	197	13	24	65	.001	24.1
	LV	74	22	9	51		
	EE	197	13	24	65	.001	18.5
	IT	83	23	19	33		
	EE	197	13	24	65	.001	41.0
	FI	66	30	14	57		
	EE	197	13	24	65	.001	44.8
	SE	55	23	19	62		
	EE	197	13	24	65	.001	34.0
	DK	60	13	19	67		
	EE	197	13	24	65	.001	47.1
	NO	50	14	18	72		
	EE	197	13	24	65	.001	45.7
	RU	57	22	12	69		
	LV	74	22	9	51	.047	7.5
	IT	83	23	19	33		
	LV	74	22	9	51	.434	2.7
	FI	66	30	14	57		
	LV	74	22	9	51	.059	7.4
	SE	55	23	19	62		
LV	74	22	9	51	.023	9.5	
DK	60	13	19	67			
LV	74	22	9	51	.005	13.0	
NO	50	14	18	72			
LV	74	22	9	51	.152	5.3	
RU	57	22	12	69			

Table 5. Continuation

Target Groups	Joy	Anger	Sadness	Neu- tral	p	Pearson's chi- square statistic as distance measure
IT	83	23	19	33	.021	9.8
FI	66	30	14	57		
IT	83	23	19	33	.001	14.5
SE	55	23	19	62		
IT	83	23	19	33	.001	18.0
DK	60	13	19	67		
IT	83	23	19	33	.001	24.8
NO	50	14	18	72		
IT	83	23	19	33	.001	19.1
RU	57	22	12	69		
FI	66	30	14	57	.440	2.7
SE	55	23	19	62		
FI	66	30	14	57	.039	8.4
DK	60	13	19	67		
FI	66	30	14	57	.020	9.8
NO	50	14	18	72		
FI	66	30	14	57	.358	3.0
RU	57	22	12	69		
SE	55	23	19	62	.363	3.2
DK	60	13	19	67		
SE	55	23	19	62	.373	3.1
NO	50	14	18	72		
SE	55	23	19	62	.570	2.0
RU	57	22	12	69		
DK	60	13	19	67	.783	1.1
NO	50	14	18	72		
DK	60	13	19	67	.262	4.0
RU	57	22	12	69		
NO	50	14	18	72	.335	3.4
RU	57	22	12	69		

Note. Pearson's χ^2 results: Cross-group difference in the recognition of an emotion is significant if $p < .05$. The statistic indicates the measure of the difference of recognition (group distance).

Table 5 demonstrates that all non-Estonian groups recognized Estonian joy significantly differently from the Estonian participants ($p < .001$). According to the statistic used, difference of recognition (i.e. distance) from the Estonian group was as follows: Italians (18.5) < Latvians (24.1) < Danes (34.0) < Finns (42.0) < Swedes (44.8) < Russians (45.7) < Norwegians (47.1). The recognition (both correct and incorrect choices) did not differ significantly, on the one hand, for Latvians, Finns, Russians and Swedes and, on the other hand, for Russians, Swedes, Danes and Norwegians, see Figure 6.

	LV	FI	RU	SE	DK	NO
LV						
FI						
RU						
SE						
DK						
NO						

Figure 6. The filled cells indicate which ethnic groups did not differ significantly in the recognition of Estonian joy.

Researching the influence of a genealogical relation of languages on the recognition of vocal emotions, we can see that the Estonian and Finnish groups, which both belong to the Finnic language branch, differed significantly in the recognition of Estonian joy. Swedish, Danish and Norwegian are also genealogically related, belonging to the North Germanic branch. Indeed, those three groups showed no significant difference in the recognition of Estonian joy, however nor did the Russian group, who belong to the East Slavic branch. Hence we can conclude that genealogical relation of languages is not decisive for similar recognition of vocal emotions.

As the genealogical relation of languages need not mean cultural similarity, a study of possible cultural influence on emotion recognition was necessary. Since all groups differed significantly from the Estonians in recognizing Estonian joy, cultural similarity with Estonians along Hofstede's four dimensions cannot play a role in emotion recognition.

Table 6. Target emotion ANGER. Comparison of test groups

Target Groups	Joy	Anger	Sadness	Neutral	p	Pearson's chi-square statistic as distance measure	
ANGER	EE	8	170	50	70	.001	20.5
	LV	15	60	26	52		
	EE	8	170	50	70	.001	21.4
	IT	19	80	14	44		
	EE	8	170	50	70	.001	32.1
	FI	24	70	17	53		
	EE	8	170	50	70	.001	53.2
	SE	28	52	17	62		
	EE	8	170	50	70	.001	63.5
	DK	32	54	11	62		
	EE	8	170	50	70	.001	28.6
	NO	16	56	21	59		
	EE	8	170	50	70	.001	20.3
	RU	57	22	12	69		
	LV	15	60	26	52	.056	7.5
	IT	19	80	14	44		
	LV	15	60	26	52	.225	4.4
	FI	24	70	17	53		
	LV	15	60	26	52	.067	7.2
	SE	28	52	17	62		
LV	15	60	26	52	.004	13.3	
DK	32	54	11	62			
LV	15	60	26	52	.767	1.1	
NO	16	56	21	59			
LV	15	60	26	52	.803	1.0	
RU	57	22	12	69			

Table 6. Continuation

Target Groups	Joy	Anger	Sadness	Neutral	p	Pearson's chi-square statistic as distance measure
IT	19	80	14	44	.528	2.2
FI	24	70	17	53		
IT	19	80	14	44	.011	11.0
SE	28	52	17	62		
IT	19	80	14	44	.008	11.8
DK	32	54	11	62		
IT	19	80	14	44	.046	8.0
NO	16	56	21	59		
IT	19	80	14	44	.281	3.8
RU	57	22	12	69		
FI	24	70	17	53	.309	3.6
SE	28	52	17	62		
FI	24	70	17	53	.163	5.1
DK	32	54	11	62		
FI	24	70	17	53	.328	3.5
NO	16	56	21	59		
FI	24	70	17	53	.564	2.0
RU	16	68	21	55		
SE	28	52	17	62	.661	1.6
DK	32	54	11	62		
SE	28	52	17	62	.288	3.8
NO	16	56	21	59		
SE	28	52	17	62	.100	6.2
RU	57	22	12	69		
DK	32	54	11	62	.038	8.4
NO	16	56	21	59		
DK	32	54	11	62	.015	10.5
RU	57	22	12	69		
NO	16	56	21	59	.778	1.1
RU	57	22	12	69		

Note. Pearson's χ^2 results: Cross-group difference in the recognition of an emotion is significant if $p < .05$. The statistic indicates the measure of the difference of recognition (group distance).

According to Table 6, all groups recognized Estonian anger significantly differently from Estonians ($p < .001$). Based on Pearson's chi-square statistics, the difference of recognition of anger (distance) from that of the Estonian group is as follows: Russians (20.3) < Latvians (20.5) < Italians (21.4) < Norwegians (28.6) < Finns (32.1) < Swedes (53.2) < Danes (63.5). The recognition (both correct and incorrect choices) did not differ significantly between Italian, Latvian, Finnish and Russian groups; or between the Latvian, Finnish, Russian, Swedish and Norwegian groups; or between the Finnish, Danish and Swedish groups, see Figure 7.

	IT	LV	FI	RU	SE	NO	DK
IT							
LV							
FI							
RU							
SE							
NO							
DK							

Figure 7. The filled cells indicate which ethnic groups did not differ significantly in the recognition of Estonian anger.

Researching the influence of genealogical relation of languages on vocal emotion recognition it was revealed that there was a significant difference between how Estonian anger was recognized by Estonian and Finnish groups, despite their belonging to the same Finnic branch. The Finns confused Estonian anger with neutral speech. In the case of the Swedish, Danish and Norwegian groups, who all speak languages of the North Germanic branch, a significant difference in the recognition of Estonian anger was observed between the Danish and Norwegian groups. At the same time, Swedish and Norwegian groups did not significantly differ in their recognition from those of the Latvian (Baltic branch), Finnish (Finnic branch) and Russian (East Slavic branch) groups. The results showed that genealogical relation of languages is not the factor behind similar recognition of a vocal emotion, either in the case of related languages or in more distant ones.

As all groups remained significantly different from Estonians in recognizing Estonian anger, recognition success could not be affected by cultural similarity in terms of Hofstede's four dimensions.

Table 7. Target emotion SADNESS. Comparison of test groups

Target Groups	Joy	Anger	Sad- ness	Neutral	p	Pearson's chi square statistic as distance measure	
SADNESS	EE	4	37	234	22	.012	11.0
	LV	3	12	115	26		
	EE	4	37	234	22	.303	3.7
	IT	3	11	131	14		
	EE	4	37	234	22	.001	15.7
	FI	9	25	108	25		
	EE	4	37	234	22	.002	15.1
	SE	3	3	138	16		
	EE	4	37	234	22	.015	10.5
	DK	1	9	126	23		
	EE	4	37	234	22	.059	7.4
	NO	7	12	118	16		
	EE	4	37	234	22	.001	16.0
	RU	10	11	117	22		
	LV	3	12	115	26	.199	4.7
	IT	3	11	131	14		
	LV	3	12	115	26	.059	7.4
	FI	9	25	108	25		
	LV	3	12	115	26	.020	9.8
	SE	3	3	138	16		
LV	3	12	115	26	.555	2.1	
DK	1	9	126	23			
LV	3	12	115	26	.262	4.0	
NO	7	12	118	16			
LV	3	12	115	26	.250	4.1	
RU	10	11	117	22			

Table 7. Continuation

Target Groups	Joy	Anger	Sad- ness	Neutral	p	Pearson's chi square statistic as distance measure
IT	3	11	131	14	.004	13.6
FI	9	25	108	25		
IT	3	11	131	14	.181	4.9
SE	3	3	138	16		
IT	3	11	131	14	.322	3.5
DK	1	9	126	23		
IT	3	11	131	14	.505	2.3
NO	7	12	118	16		
IT	3	11	131	14	.096	6.3
RU	10	11	117	22		
FI	9	25	108	25	.001	25.8
SE	3	3	138	16		
FI	9	25	108	25	.002	15.2
DK	1	9	126	23		
FI	9	25	108	25	.084	6.6
NO	7	12	118	16		
FI	9	25	108	25	.117	5.9
RU	10	11	117	22		
SE	3	3	138	16	.122	5.8
DK	1	9	126	23		
SE	3	3	138	16	.038	8.4
NO	7	12	118	16		
SE	3	3	138	16	.012	11.0
RU	10	11	117	22		
DK	1	9	126	23	.096	6.3
NO	7	12	118	16		
DK	1	9	126	23	.048	7.9
RU	10	11	117	22		
NO	7	12	118	16	.713	1.4
RU	10	11	117	22		

Note. Pearson's χ^2 results: Cross-group difference in the recognition of an emotion is significant if $p < .05$. The statistic indicates the measure of the difference of recognition (group distance).

Table 7 shows that for recognition of Estonian sadness, the Italian group ($p=.303$) and Norwegian group ($p=.059$) did not differ significantly from the Estonian group. Based on Pearson's chi-square statistics, the difference of recognition of sadness (distance) from that of the Estonian group is as follows: Italian (3.7) < Norwegian (7.4) < Danish (10.5) < Latvian (11.0) < Swedish (15.1) < Finnish (15.7) < Russian (16.0). The recognition (both correct and incorrect choices) did not differ significantly between Estonian, Italian and Norwegian groups; or between those of the Swedish Italian and Danish groups; or between the Italian, Norwegian, Danish and Latvian groups; or between the Norwegian, Latvian, Finnish and Russian groups, see Figure 8.

	EE	SE	IT	DK	NO	LV	FI	RU
EE								
SE								
IT								
DK								
NO								
LV								
FI								
RU								

Figure 8. The filled cells indicate which ethnic groups did not differ significantly in the recognition of Estonian sadness.

To reveal the possible effect of genealogical relation of languages we looked into groups that spoke languages belonging to Finnic and North Germanic branches. The Finnish group recognized Estonian sadness significantly differently from the Estonian group. In the North Germanic branch (Norwegian, Swedish, Danish), a significant difference in the recognition of Estonian sadness was observed between the Swedish and Norwegian groups. In the North Germanic branch, the Danish group did not differ significantly from either the Norwegian or the Swedish groups in their recognition of Estonian sadness, but a significant difference was observed between the Swedish and Danish groups. Consequently, the belonging to one and the same language branch is not a decisive factor

in similar recognition of a vocal emotion expressed either in a related language or in a more distant one.

Estonian sadness was recognized well by all groups, but there were only two groups – Italian and Norwegian – whose recognition of sadness (including both correct and incorrect choices) did not differ significantly from the Estonian group. Taking into account the cultural distances (based on Hofstede’s dimensions) between the Estonians and the rest of the groups, notably,

Individualism:

EE < FI (7)³ < NO (9) < LV (10) < SE (11) < DK (14) < IT (16) < RU (21)

Power distance:

EE < LV (4) < FI (8) > SE (9) = NO (9) < IT (10) < DK (22) < RU (53)

Uncertainty avoidance:

EE < FI (1) < LV (3) < NO (10) < IT (15) < SE (31) < RU (35) < DK (37)

Masculinity:

EE < FI (4) < RU (6) < DK (14) < LV (21) < NO (22) < SE (21) < IT (40),

we can see that there are cultures whose values are closer to Estonian ones than Italian and Norwegian cultures, and yet their recognition of Estonian sadness was significantly different from that of the Estonian group. Consequently, no relation is observed between closer cultural distance (in Hofstede’s terms) and recognition similarity.

According to Table 8 the recognition of Estonian neutral speech by the Swedish ($p=.172$), Danish ($p=.130$), and Norwegian ($p=.057$) groups did not differ significantly (in both correct and incorrect choices) from the Estonian group. Based on Pearson’s chi-square statistics, the difference of recognition of neutrality (distance) from that of the Estonian group is as follows: Swedish (5.0) < Danish (5.6) < Norwegian (7.5) < Latvian (9.4) < Italian (24.2) < Russian (25.4) < Finnish (34.6). There was no significant difference of recognition between the Estonian, Swedish, Danish and Norwegian groups; or between the Swedish, Danish, Norwegian and Latvian groups; or between the Norwegian, Latvian and Russian groups; or between the Russian and Italian groups; or between the Italian and Finnish groups, see Figure 9.

3 Parentheses contain distance from Estonian culture, see Table 2 for indices.

Table 8. Target NEUTRAL SPEECH. Comparison of test groups

Target Groups	Joy	Anger	Sadness	Neutral	p	Pearson's chi-square statistic as distance measure	
NEUTRAL	EE	10	16	17	107	.025	9.4
	LV	9	18	11	42		
	EE	10	16	17	107	.001	24.2
	IT	23	12	5	39		
	EE	10	16	17	107	.001	34.6
	FI	29	13	4	38		
	EE	10	16	17	107	.172	5.0
	SE	4	16	11	47		
	EE	10	16	17	107	.130	5.6
	DK	7	12	16	45		
	EE	10	16	17	107	.057	7.5
	NO	6	19	8	46		
	EE	10	16	17	107	.001	25.4
	RU	15	24	6	35		
	LV	9	18	11	42	.021	9.7
	IT	23	12	5	39		
	LV	9	18	11	42	.002	14.7
	FI	29	13	4	38		
	LV	9	18	11	42	.513	2.3
	SE	4	16	11	47		
LV	9	18	11	42	.479	2.5	
DK	7	12	16	45			
LV	9	18	11	42	.734	1.3	
NO	6	19	8	46			
LV	9	18	11	42	.216	4.5	
RU	15	24	6	35			

Table 8. Continuation

Target Groups	Joy	Anger	Sadness	Neutral	p	Pearson's chi-square statistic as distance measure
IT	23	12	5	39	.872	0.7
FI	29	13	4	38		
IT	23	12	5	39	.001	16.9
SE	4	16	11	47		
IT	23	12	5	39	.002	14.7
DK	7	12	16	45		
IT	23	12	5	39	.005	12.8
NO	6	19	8	46		
IT	23	12	5	39	.112	6.0
RU	15	24	6	35		
FI	29	13	4	38	.001	23.3
SE	4	16	11	47		
FI	29	13	4	38	.001	21.2
DK	7	12	16	45		
FI	29	13	4	38	.001	18.2
NO	6	19	8	46		
FI	29	13	4	38	.043	8.2
RU	15	24	6	35		
SE	4	16	11	47	.506	2.3
DK	7	12	16	45		
SE	4	16	11	47	.769	1.1
NO	6	19	8	46		
SE	4	16	11	47	.011	11.2
RU	15	24	6	35		
DK	7	12	16	45	.228	4.3
NO	6	19	8	46		
DK	7	12	16	45	.005	12.7
RU	15	24	6	35		
NO	6	19	8	46	.102	6.2
RU	15	24	6	35		

Note. Pearson's χ^2 results: Cross-group difference in the recognition of an emotion is significant if $p < .05$. The statistic indicates the measure of the difference of recognition (group distance).

	EE	SE	DK	NO	LV	RU	IT	FI
EE								
SE								
DK								
NO								
LV								
RU								
IT								
FI								

Figure 9. The filled cells indicate which ethnic groups did not differ significantly in the recognition of Estonian neutral speech.

The influence of genealogical relation of languages on the recognition of neutral speech shows opposite tendencies: Notably, the Finns, who belong to the same language (Finnic) branch as Estonians, differed significantly in the recognition of Estonian neutrality. The three groups belonging to the North Germanic branch, the Swedish, Danish and Norwegian groups, did not differ significantly in their recognition of Estonian neutrality, either from the Estonian group or from each other. At the same time, there was no significant difference between those three and the Latvians, who belong to the Baltic branch, yet the Latvian group differed significantly from the Estonian group in recognizing neutral speech.

An analysis of cultural influence on neutrality recognition reveals that Estonian neutral speech was recognized correctly by the Swedish, Danish and Norwegian groups, whereas the rest would confuse it with other emotions. In Hofstede's terms, the cultural backgrounds of the Swedes, Danes and Norwegians share individualism, low power distance and femininity, with uncertainty being the only distinctive dimension, as Norwegians avoid uncertainty more than Swedes or Danes. This means these cultures are comparatively similar. However, Estonians are culturally closer to Finns and Latvians, who nevertheless differed significantly from Estonians in recognizing Estonian neutral speech (the Finnish group confused it with joy). Thus, it is not possible to say unequivocally that the genealogical relation of languages or

cultural closeness in terms of Hofstede's four dimensions are decisive factors in the recognition of vocal neutrality in another language.

Finally, the distance between all the test groups was calculated taking into account all emotions and their classifications (see Table 9).

Table 9. Pearson's chi-square statistic as distance measure of emotion recognition

	LV	IT	FI	SE	DK	NO	RU
EE	64.9	67.8	123.4	118.1	113.6	90.7	107.4
LV		29.8	29.3	26.7	27.4	19.4	14.9
IT			26.3	47.3	48.0	48.0	35.3
FI				55.4	49.9	38.0	19.1
SE					12.9	16.4	30.4
DK						20.2	35.1
NO							12.1

If we rank the groups according to their similarity to the Estonians in emotion recognition, it is the Latvians who are the closest to the Estonians. after that comes the Italian group, while the Finns classify Estonian emotions the most differently: **EE** < LV < IT < NO < RU < DK < SE < FI.

As to cultural similarity in terms of Hofstede's four dimensions, it obviously did not affect the recognition of Estonian vocal emotions. Within the eight groups analysed, Estonia was culturally the most similar with Finland and Latvia, and the most different from Russia: **EE** < FI < LV < NO < SE < IT < DK < RU (see Table 3).

Nor does a comparison of the general performance in the recognition of Estonian vocal emotions (Table 9) reveal any pattern following either genealogical relation of languages or cultural closeness. The patterns observed are either random or depending on factors beyond our analysis.

4. Discussion

The aim of the study was to test hypotheses of universality and similarity: to see whether cultures that speak different languages and have different cultural values will be able to recognize vocal emotions

with an accuracy better than chance probability without confusing them with other emotions; and to verify whether cultures with similar cultural values and of the same language branch perform similarly in emotion recognition.

A comparison of the performance of test groups with different cultural values and of different languages – Estonians, Latvians, Finns, Swedes, Danes, Norwegians, Italians and Russians – in the recognition of Estonian vocal emotions partially confirmed the universality hypothesis, whereas the similarity hypothesis was not confirmed. Although Estonian vocal emotions were recognized better than chance probability by all groups, the target emotion was often confused with other emotions (see Table 4). All emotions were recognized correctly, that is, without confusion, by the Estonian group, which is suggestive of in-group advantage. The only group of those who did not speak Estonian and did not live in Estonia, yet recognized Estonian joy was the Italian group. The rest of the non-Estonian groups tended to confuse Estonian joy with neutrality. Similarly, Estonian anger was confused with neutrality. Estonian neutrality was recognized correctly by the Latvians, Swedes, Danes, and Norwegians, whereas the Russian group confused it with anger, the Finns and Italians with joy. Estonian sadness, however, was recognized by all test groups, without confusion with other emotions or neutrality (see Table 10). According to these results only sadness can be regarded as universally recognizable as it was not confused by any group with any other option.

Several cross-culture studies have shown better recognition for negative emotions like sadness or anger (e.g. Jürgens et al. 2013, Pell et al. 2009, Scherer 2011, and Thompson and Balkwill 2006). This is in accord with the fact that Estonian sadness was recognized relatively well by all test groups. However, the non-Estonian groups failed to recognize Estonian anger, which was confused with neutral speech (cf. Altrov 2013). Consequently, our results do not confirm that different culture groups would recognize moderately expressed elicited emotions any better if the emotions were negative.

Table 10. Recognition of Estonian emotions by Latvians, Italians, Finns, Swedes, Danes, Norwegians and Russians (% of target recognition)

Groups	Target emotion			
	Joy	Anger	Sadness	Neutral
Estonians	65.9	57.0	78.8	71.3
Latvians			73.7	52.5
Italians	52.5		82.4	
Finns			64.7	
Swedes			86.3	60.3
Danes			79.2	56.3
Norwegians			77.1	58.2
Russians			73.1	

The fact that Estonian vocal emotions were recognized with excess probability and part of them were not confused with other emotions does not mean that the recognition of vocal emotions by non-Estonian groups was similar to those of Estonians. As can be seen from Tables 5–9, most of the test groups differed significantly from Estonians in the recognition of most emotions. Estonian joy and anger were recognized significantly differently from Estonians by all non-Estonian groups; for Estonian sadness there was no significant difference between Estonians, Italians and Norwegians, while for neutral speech there was no significant difference between Estonians, Swedes, Danes and Norwegians (see Table 11).

Table 11. Similarity (marked with an x) between Estonians and the other test groups in the recognition of target emotions

Groups	Target emotion			
	Joy	Anger	Sadness	Neutral
Latvians				
Italians			x	
Finns				
Swedes				x
Danes				x
Norwegians			x	x
Russians				

To test the similarity hypothesis we investigated whether the cultures that are closer to each other, linguistically or culturally, perform similarly in emotion recognition, making similar choices between given options. In a comparison of the recognition patterns of the Norwegian, Swedish and Danish participants, who all belong to the North Germanic branch, we can observe some similarity only between the Swedish and Danish groups, whereas the Norwegian choices coincide rather more with the Russian ones (see Table 9). The influence of the genealogical relation of languages could also be studied by comparing the performance of Estonian and Finnish groups, as both languages belong to the Finnic branch of the Uralic language family. It was revealed that in comparison with other groups it was the Finns whose emotion recognition differed the most from that of the Estonians, rather resembling the Russian and Italian groups (see Table 9). Nor did genealogical relation of language give any advantage to the Finns in correct recognition of Estonian vocal emotions: the only Estonian emotion they did not confuse with any other emotion was sadness (see Table 10).

Based on these results, we cannot unambiguously conclude that speakers of quite closely related languages would recognize emotions similarly (examples: Swedish, Danish and Norwegian; or Estonian and Finnish) or whether genealogical relation of languages provides an advantage in recognizing the emotions expressed in a related language (example: Finnish). In order to assess the cultural distance of the countries represented by the test groups we used Hofstede's indices for four cultural dimensions (see Hofstede 2001, and Hofstede et al. 2010) – individualism, power distance, uncertainty avoidance, and masculinity. Using those four indices we determined which of the cultures involved were closer (more similar) to the Estonian culture and to one another, and investigated whether similar cultural values were a factor in the recognition of Estonian vocal emotions. According to Table 3, Estonian cultural values are the closest to those of Finland and Latvia, while Russia is culturally the most distant country from Estonia. As to emotion recognition, the Latvian group was the most similar to the Estonian one, but the Finnish group was the most dissimilar (see Table 9). Hence we cannot confirm that cultural closeness is a positive factor in emotion recognition.

Earlier cross-cultural studies have partially confirmed the hypothesis that collectivist cultures can be better in recognizing negative emotions than individualistic cultures (see Jürgens 2013). According to our results, however, the Russians, who represent a collectivist culture,

were no better in recognizing negative emotions than other test groups: they were as likely to recognize Estonian sadness as the test groups from individualistic cultures, and they did not recognize Estonian anger.

Consequently, according to our results neither genealogical relation of languages nor cultural similarity play an important role in the recognition of emotions expressed in another language (no in-group advantages) and this is where our results differ from those of several previous studies (cf. Paulmann and Uskul 2014, Pell et al. 2009, and Scherer et al. 2001).

Our results suggest that the vocal everyday expression of moderate emotions differs across cultures, while emotions heard in other languages are usually interpreted on the basis of their acoustic similarity to the typical emotional expression of the listener's native language. For example, knowing that the Italians were the only group who recognized Estonian joy, we can assume that Italian vocal joy can also be characterized by a high pitch, average intensity and average speech rate. Or again, if Russians confused Estonian anger with neutrality, we can assume that Russian neutral speech has a low pitch, average intensity and high speech rate (cf. Altrov 2013). To what extent those assumptions hold needs testing by a comparison of the acoustic characteristics of comparable samples of the relevant languages. The available material is too different from ours to enable a comparative approach. Notably, we had the test groups listen to moderately expressed emotions of ordinary people, whereas many earlier studies were based on pseudo-sentences instructed to be read by actors expressing various emotions. It is likely that emotions acted out by professional actors are easier to recognize cross-culturally, as such play-acted expression is, as a rule, stereotypical and exaggerated.

According to the literature, the emotions of another culture are difficult to recognize merely by ear, without seeing the facial expression. Our results demonstrated that Estonian vocal joy, anger and neutrality were often misinterpreted by groups representing other cultures, which indicates that the acoustic expression of those emotions may easily be culturally specific. As to the recognition rate of Estonian sadness it was very high among all of the groups, which supports the view that the acoustic expression of sadness can be universal, being characterized by a slow speech rate and low intensity.

5. Summary

Expression as well as understanding of emotional speech involves both universal and language- or culture-specific aspects. The aim of our study was to find out to what extent the understanding of vocal emotions could depend on linguistic or cultural affiliation. The research questions were: (1) Do people from different cultural and linguistic backgrounds identify emotions accurately, without confusing the target emotion with other emotions? (2) Do the representatives of cultures with similar values or of the same language family identify emotions similarly?

To answer those questions we studied the audio recognition of moderately expressed elicited emotions in Estonian speech by people not living in Estonia and not speaking Estonian, who represented different language families and different cultures. Besides Estonians, the participants of the experiment included groups of Latvians, Danes, Norwegians, Swedes, Finns, Italians and Russians, all living in their titular countries. None of the foreign groups spoke Estonian. Apart from the linguistic difference, the groups could be classified under different cultures according to Hofstede (2001).

The participants were asked to listen to single out-of-context sentences without seeing the text and decide upon the emotion of each sentence. The choice was between anger, joy, sadness and neutral.

All test groups recognized the sentence emotion with excess probability, but only the Estonian group was homogeneous in their decisions. Notably, part of the Latvians, Swedes, Danes and Norwegians mistook Estonian joy and anger for neutral; the Italians confused Estonian anger with neutral and Estonian neutrality with joy; the Russians confused Estonian joy and anger with neutral and Estonian neutrality with anger. Estonian joy was recognized correctly by Italians only, while Estonian neutrality was recognized correctly by the Latvian, Swedish, Danish and Norwegian groups. Estonian sadness, however, was recognized by all groups without confusion with any other emotion. Consequently, according to our results only sadness can be universally recognized.

Our comparison of groups speaking genealogically related languages as well as of those with a similar cultural background did not show similarities in the recognition of vocal emotions by related languages, nor was there an advantage recognizing emotions in genealogically related languages. Neither did our results demonstrate any advantage of cultural similarity in emotion recognition.

Acknowledgements

This work was supported by institutional research funding IUT35-1 of the Estonian Ministry of Education and Research.

Addresses:

Rene Altrov
Institute of the Estonian Language
Roosikrantsi 6
10119 Tallinn, Estonia
E-mail: rene.altrov@eki.ee

Hille Pajupuu
Institute of the Estonian Language
Roosikrantsi 6
10119 Tallinn, Estonia
E-mail: hille.pajupuu@eki.ee

References

- Altrov, Rene (2013) “Aspects of cultural communication in recognizing emotions”. *Trames* 17, 159–174.
- Altrov, Rene and Hille Pajupuu (2010) “Estonian Emotional Speech Corpus: culture and age in selecting corpus testers”. In Inguna Skadiņa and Andrejs Vasiljevs, eds. *Human language technologies – the Baltic perspective – proceedings of the fourth international conference Baltic HLT 2010*, 25–32. Amsterdam: IOS Press.
- Altrov, Rene and Hille Pajupuu (2012) “Estonian Emotional Speech Corpus: theoretical base and implementation”. In Laurence Devillers, Björn Schuller, Anton Batliner, Paolo Rosso, Ellen Douglas-Cowie, Roddy Cowie, and Catherine Pelachaud, eds. *4th international workshop on corpora for research on emotion sentiment & social signals (ES3)*, 50–53. Istanbul.
- Bryant, Gregory A. and Clark H. Barrett (2008) “Vocal emotion recognition across disparate cultures”. *Journal of Cognition and Culture* 8, 1/2, 135–148.
- Cowie, Roddy, Naomi Sussman, and Aaron Ben-Ze’ev (2011) “Emotions: concepts and definitions”. In Paolo Petta, Catherine Pelachaud, and Roddy Cowie, eds. *Emotion-oriented systems: the HUMAINE handbook*, 9–31. Berlin, Heidelberg: Springer.
- Elfenbein, Hillary A. (2013) “Nonverbal dialects and accents in facial expressions of emotion”. *Emotion Review* 5, 90–96.
- Hofstede, Geert (2001) *Culture’s consequences: comparing values, behaviors, institutions, and organizations across nations*. 2nd ed. Thousand Oaks, CA: Sage Publications.

- Hofstede, Geert, Gert Jan Hofstede, and Michael Minkov (2010) *Cultures and organizations: software of the mind*. Revised and expanded 3rd ed. New York: McGraw-Hill USA.
- Jürgens, Rebecca, Matthis Drolet, Ralph Pirow, Elisabeth Scheiner, and Julia Fischer (2013) "Encoding conditions affect recognition of vocally expressed emotions across cultures". *Frontiers in Psychology. Emotion Science* 4, 111, 1–10.
- Kamaruddin, Norhaslinda, Abdul Wahab, and Chai Quek (2012) "Cultural dependency analysis for understanding speech emotion". *Expert Systems with Applications* 39, 5115–5133.
- Laukka, Petri, Daniel Neiberg, and Hillary A. Effenbein (2014) "Evidence for cultural dialects in vocal emotion expression: acoustic classification within and across five nations". *Emotion* 13, 3, 445–449.
- Paulmann, Silke and Ayse K. Uskul (2014) "Cross-cultural emotional prosody recognition: evidence from Chinese and British listeners". *Cognition and Emotion* 28, 2, 230–244.
- Pell, Marc D., Laura Monetta, Silke Paulmann, and Sonja A. Kotz (2009) "Recognizing emotions in a foreign language". *Nonverbal Behaviour* 33, 107–120.
- Scherer, Klaus R., Rainer Banse, and Harald G. Wallbott (2001) "Emotion inferences from vocal expression correlate across languages and cultures". *Journal of Cross-Cultural Psychology* 32, 76–92.
- Scherer, Klaus R., Elizabeth Clark-Polner, and Marcello Mortillaro (2011) "In the eye of the beholder? Universality and cultural specificity in the expression and perception of emotion". *International Journal of Psychology* 46, 6, 401–435.
- Tamuri, Kairi (2012) "Intensity of Estonian emotional speech". Human language technologies – the Baltic perspective – proceedings of the fifth international conference Baltic HLT 2012, 238–246. Amsterdam: IOS Press.
- Tamuri, Kairi and Meelis Mihkla (2012) "Emotions and speech temporal structure". *Linguistica Uralica* 3, 209–217.
- Tamuri, Kairi (2015) "Fundamental frequency in Estonian emotional read-out speech". *Journal of Estonian and Finno-Ugric Linguistics* 6, 1, 9–21.
- Thompson, William F. and L.-L. Balkwill (2006) "Decoding speech prosody in five languages". *Semiotica* 158, 407–424.

Kokkuvõte. Rene Altrov ja Hille Pajupuu: Keele ja kultuuri mõju kõneemotsioonidest arusaamisele. Emotsionaalse kõne väljendamises ja selle mõistmises on nii universaalseid kui ka kultuuri- ja keelespetsiifilisi aspekte. Meie uurimuse eesmärk oli teada saada, kuidas mõjutab keeleline ja kultuuri-line kuulumine kõneemotsioonidest arusaamist: esiteks, kas eri kultuuridest ja eri keeli kõnelevad inimesed tunnevad emotsioone ühte viisi ära ega aja neid segi teiste emotsioonidega, ja teiseks, kas sarnastest kultuuridest ja samasse keelerühma kuuluvad inimesed tunnevad emotsioone ära sarnaselt. Selleks uurisime eesti keeles mõõdukalt väljendatud esilekutsutud emotsioonide äratundmist Eestis mitteelavate ja eesti keelt mitteoskavate inimeste poolt, kes

valiti erinevatest keelerühmadest ja kultuuridest. Osalejatena kaasati lisaks eestlastele kui korpuseandmeid kinnitavale kontrollrühmale lätlased, taanlased, norrakad, rootslased, soomlased, itaallased ja venelased, kes elasid oma emamaal. Peale eestlaste ei osanud ükski testirühm eesti keelt. Osalejatel palusime kuulata Eesti emotsionaalse kõne korpuse kontekstita üksiklauseid, ilma et nad teksti oleks näinud, ja otsustada, millise emotsiooniga iga lause puhul tegu on. Valida sai viha, rõõmu, kurbuse ja neutraalsuse vahel. Kõik testirühmad tundsid lause emotsiooni ära üle juhusliku valiku tõenäosuse, kuid erinevalt eestlastest ei olnud teised testirühmad emotsioonide tuvastamisel ühtsed. Osa lätlasi, rootslasi, taanlasi ja norrakaid pidas eesti rõõmu ja viha neutraalseks; itaallased ajasid viha segi neutraalsusega ja neutraalse rõõmuga; venelased ajasid rõõmu ja viha segi neutraalsusega ja neutraalse vihaga. Eesti rõõmu tundsid õigesti ära vaid itaallased ja eesti neutraalsuse lätlased, rootslased, taanlased ja norrakad. Eesti kurbuse tundsid seevastu ära kõik testirühmad ja seda ei aetud segi muude emotsioonidega. Seega, meie uurimuse tulemuste põhjal saab universaalselt äratuntud emotsiooniks pidada vaid kurbust. Võrreldes sugulaskeeli ja sarnaseid kultuure omavahel, ei näidanud meie uuringu tulemused, et keelesugulased tuvastaks kuuldud emotsioone ühtmoodi või et keelesugulus annaks sugulaskeelte emotsioonide äratundmisel eelise. Samuti ei saa meie uurimuse tulemuste põhjal kinnitada, et kultuuride sarnasus aitaks kaasa emotsioonide tuvastamisele.

Märksõnad: emotsioonid, emotsionaalne kõne, tajut, kultuuridevaheline, sugulaskeeled, Euroopa keeled